NATIONAL COMMISSION ON NEW TECHNOLOGICAL USES OF COPYRIGHTED WORKS
(CONTU)

Tel: (202) 557-0996 Washington, D.C. 20558

DATA BASE SUBCOMMITTEE REPORT
&
ADDITIONAL VIEWS

The National Commission on New Technological Uses of Copyrighted Works, in compliance with that portion of its mandate which instructs it to examine questions concerning computer uses of copyrighted works, has designated a Data Base Subcommittee to prepare a preliminary report concerning the manner in which automated data bases should be dealt with by copyright. That report, containing the recommendations of the Data Base Subcommittee, is herewith presented for public comment. It has not been adopted by the full Commission, which will base its final report to the President and the Congress upon response to this document, further hearings, and the results of several studies now in progress concerning the economic, public interest, and consumer questions associated with the issues before the Commission.

Appended to the report are the individual views of those Commissioners who desire at this time to comment upon the Subcommittee's recommendations.

Written comments or requests for the opportunity to testify on this subject at future Commission meetings should be addressed to Arthur J. Levine, Executive Director, National Commission on New Technological Uses of Copyrighted Works, Washington, D.C. 20558. Requests to appear must be received by August 1, 1977, so that the Commission may plan appropriate hearings early in the fall. Written comments would be welcome at any time but should be submitted by September 1, 1977.

NATIONAL COMMISSION ON NEW TECHNOLOGICAL USES OF COPYRIGHTED WORKS
(CONTU)

Tel: (202) 557-0996 Washington, D.C. 20559

REPORT OF THE DATA BASE SUBCOMMITTEE TO THE NATIONAL COMMISSION ON NEW TECHNOLOGICAL USES OF COPYRIGHTED WORKS

The automated data base represents a new technological use of a type of work long recognized as eligible for copyright. Dictionaries, encyclopedias, and tables of numeric information are all forms of data bases which antedate the computer by many decades, and for which copyright protection has been, and will continue to be, available under applicable copyright law. Under the new law a data base is a compilation and thus a proper subject for copyright.[1/] This entitlement to copyright is not diminished by the fixation of the information content of a data base in a medium requiring the intervention of a computer to accomplish the communication of content.[2/] Accordingly a data base, whether printed in traditional hard copy or fixed

---

1/ Section 101, P.L. 94-553 defines "compilation" as

> a work formed by the collection and assembling of pre-existing materials or of data that are selected, coordinated, or arranged in such a way that the resulting work as a whole constitutes an original work of authorship. The term "compilation" includes collective works.

2/ Section 102(a), P.L. 94-553, provides that

> Copyright protection subsists, in accordance with this title, in original works of authorship fixed in any tangible medium of expression, now known or later developed, from which they can be perceived, reproduced, or otherwise communicated, either directly or with the aid of a machine or device (emphasis added).

on electromagnetic tape, is protected by copyright under the terms of the new law. 3/

Computer-readable data bases do differ, of course, from their hard-copy counterparts. Some of these differences raise copyright issues and related policy considerations, and the Data Base Subcommittee has attempted to identity those societal interests and values which could be advanced by the Commission's recommendations with regard to data bases. Copyright applied to data bases should, it was generally agreed, encourage the development and dissemination of useful stores of information, so as to make this information readily availabe to the public. In addition, the Commission should encourage data base proprietors to publish and register their copyrighted works, which would create a public record of the existence of the works and, in turn, make possible public awareness and utilization of the works. 4/

3/ The House of Representatives Report accompanying P. L. 94-553 makes clear the intention to include computer-readable data bases within copyright by explaining that:

> The term 'literary works' does not connote any criterion of literary merit or qualitative value: it includes catalogs, directories, and similar factual references, or instructional works and compilations of data. It also includes computer data bases...H.R. Rep. 94-1476, 94th Cong., 2nd Sess. 54 (1976) [hereinafter cited as "H.R. 94-1476"].

4/ Maximization of public access to information contained in automated data bases is cited as a significant goal of a national information policy in the Report to the President of the United States on National Information Policy 70 (1976), prepared by the Domesitc Council Committee on the Right of Privacy, under the chairmanship of then-Vice President Nelson Rockefeller.

The following discussion is premised upon the Commission's agreement with the legislative intent of the new copyright law to accord copyright protection for computer data bases equivalent with the protection accorded compilations in traditional hard-copy format. The problem-areas identified by the Commission are: 1) What copyright consequences attach to the "input" into a computer of a copyrighted work (perhaps better described as the fixation of a work in a medium capable of use within a computer system)? 2) What rights does the proprietor of copyright in a data base have in regard to the use of extracts provided in response to authorized searches or enquiries made of the data base? and 3) What constitutes publication of a data base, and what legal consequences attach to publication? 5/

I. The "Input" Issue

The issue whether copyright liability should attach at the "input" or "output" stage of use in conjunction with a computer, i.e., at the time a work is placed in machine-readable form in a computer memory unit

---

5/ It should be clear that the same principles which apply to data bases apply also to any copyrightable works embodied in a format for use and reproduction within a computer.

or when access is sought to the work existing in computer memory, has been the primary source of disagreement regarding copyright protection for works in computer-readable form. This issue provided the major impetus for the introduction of Section 117 into the copyright revision bill, which was designed to delay reaching any final resolution of the "input-output" problem, without delaying passage of the copyright revision bill, until further consideration of the issue could be undertaken by a then still-to-be-created commission (CONTU).[6/] It appears, nevertheless

---

6/ Section 117, P.L. 94-553, provides:

> Notwithstanding the provisions of sections 106 through 116 and 118 this title does not afford the owner of copyright in a work any greater or lesser rights with respect to the use of the work in conjunction with automatic systems capable of storing, processing, receiving, or transferring information, or in conjunction with any similar device, machine, or process, than those afforded to works under the law, whether title 17 or the common law or statutes of a State, in effect on December 31, 1977, as held applicable and construed by a court in an action brought under this title.

This section was first introduced into the copyright revision bill in 1969, see S. 543, 91st Cong., 1st Sess. (1969), at which time the impact of the computer, and particularly the "input-output" question, was causing great concern on the part of copyright proprietors. Section 117 was agreed upon by interested parties as a means of permitting passage of the copyright revision without committing the Congress to a position on the computer-related issue until more study could be undertaken.

that the provisions of the new copyright law offer appropriate and sufficient guidance to determine what acts create copyright liability in this area. The protection afforded by section 106 of the act would seemingly prohibit the unauthorized storage of a work within a computer memory which, being merely one form of reproduction, would be one of the exclusive rights granted by copyright.[7/]

Considering the act of storing a computerized data base in the memory of a computer as an exclusive right of the copyright proprietor appears consistent both with accepted copyright principles and with considerations of fair treatment for potentially affected parties. Making a copy of an entire work would normally, subject to some possible exception for fair use, be considered exclusively within the domain of the copyright proprietor. One would have to assume, however, that fair use would apply rarely to the reproduction in their entirety of compendious works, such as data bases.[8/] If a copy of the work is to be stored in a computer and subsequently made accessible to others, its creation would have to be properly authorized by the copyright proprietor. The fact that only one copy is being made, or even that the owner of the computer system intends to exact no fee for providing access to the work, would no more insulate the copies from liability for copyright infringement

---

7/ It may be that the use of the term "input" to describe the act to which copyright liability attaches has been misleading. A more accurate description of the process by which a work may be stored in a computer memory would indicate that a reproduction is created within the computer memory in order to make the work accessible by means of the computer.

8/ See section 107, P. L. 94-553 for statutory criteria governing "fair use."

than would similar circumstances insulate a public library which made unauthor duplications of entire copyrighted works for its basic lending functions. 9/

Under normal circumstances, the transfer by sale or lease of a copyrighted work in computer-readable form, such as a data base, would be a meaningless transaction unless implicit in the transfer was the authorization to place or reproduce a copy in the memory unit of the transferee's computer. Any limitations on the use to be made of the copy would be a matter to be negotiated between private parties, guided by applicable public policy considerations. 10/ The proprietor of a work in computer-readable form would, under any foreseeable circumstances, be able to control by contract the future disposition of machine-readable copies of his proprietary work. Thus, the proprietor of copyright in such a work would always have a valid cause of action, arising either under copyright or contract, if a reproduction of the work is entered into a computer without the proprietor's authorization, or if a transferee authorizes a third party to enter a copy into the memory unit of a computer in violation of the terms of a valid agreement with the proprietor.

9/ The example of a copyrighted work placed in a computer memory solely to facilitate an individual's scholarly research has been cited in earlier Commission meetings as a possible fair use. The Data Base Subcommittee agrees that such a use, restricted to individual research, should be considered fair use. In order to prevent abuse of the "permission" provided under fair use principles, any "copy" created in a machine memory should be erased after completion of the particular research project for which it was made. This "copy" could be retained, for archival or further research purposes, only with authorization from the copyright proprietor.

10/ Outright sale by a copyright proprietor of a copy of a protected work, rather than a lease under which the proprietor retains ownership of a copy which the lessee may use in accord with negotiated terms and conditions, normally results in a complete loss of control over the copy which has been sold. This reflects the unwillingness of courts to enforce restrictions on the alienation of property, once a complete transfer of ownership interest in any item of property has been accomplished.

The fact that copyright would not provide the sole right and remedy for unauthorized use of a protected work is neither unique to the protection of proprietary interests in computer-readable works, nor is it a situation to be considered undesirable. 11/

Accordingly, the Data Base subcommittee believes that the application of principles already embodied in the language of the new copyright law achieves the desired substantive legal protection for copyrighted works which exist in machine-readable form. The introduction of a work into computer memory would, consistent with the new law, be a reproduction of the work which is one of the exclusive rights of the copyright proprietor. The unauthorized transfer of an existing machine-readable embodiment of a work could subject the violators to remedies for breach of contract. Principles of fair use would be applicable in limited instances to excuse an unauthorized "input" of a work into computer memory. Exemplifying such fair uses could be the creation of a copy in computer memory in order to prepare a concordance of a work, or to perform a syntactical analysis of a work, which but for the use of a computer would require a prohibitive amount of human time and effort. To satisfy the criteria of fair use, any copies created for such research purposes should be destroyed upon completion

---

11/ Remedies for breach of contract, if the right being protected is not equivalent to copyright, would not be preempted under the provisions of Section 301 of the new copyright law, and would accordingly be available to one who, on the strength of a copyright interest, granted permission to another to make certain uses of the copyrighted work only to have the terms of the authorization violated. There continues to be some scope for state enforcement of proprietary rights in intellectual property under the new copyright law. See H.R. 94-1476, supra note 3 at 131-32. The fact that state, rather than federal, law would be involved presents few real problems. The existence of parallel, but not equal rights under state and federal law reflects advantages as well as disadvantage's inherent in a federal polity, and in event both claims could be joined in the same federal cause of action under principles of pendent jurisdiction.

for the research project for which they were created. Should the individual or institution carrying on this research desire to retain the copy for archival purposes or future use, it should be required to obtain permission to do so from the copyright proprietor. All these provisions could be explained in CONTU's final report and, depending upon actions taken by Congress pursuant to the report, would be considered an interpretive aid to the copyright law akin to legislative reports.

II. Scope of Copyright in a Data Base

A computer-readable data base derives its value in large part from the ease with which a user may retrieve from it data conforming to certain specifications. That ease is the product of several factors -- the organization of the data, the sophistication of the program which assists in the searching and retrieving, and the skill of the searcher in articulating the search criteria. The difference between using a data base in hard copy and one in computer-readable form is that the former is passive and the latter may be, in the language of the industry, interactive.[12/] Thus a student who searches the Reader's Guide to Periodical Literature (a copyrighted data base) must not only know what is sought but must also painstakingly read much unsought material found in numerous volumes and updates to obtain the desired information. If, however, an interactive bibliographic data base is used only the topic(s) of interest need be expressed in order to receive citations to apparently pertinent literature and, frequently, abstracts of that literature to allow further evaluation of its utility. One important question for the Commission's purposes concerns what rights the proprietor of a computer-readable data base has in the information obtained pursuant to a user's request to, or "search" of, such a data base.

12/ An "interactive" data base is one with which a user, aided by a computer, can "converse," i.e., the user frames questions to which the data base, controlled by a computer, provides response.

There appears little doubt that one who obtained access to a copyrighted data base by normal commercial methods -- paying the proprietor or the proprietor's authorized agent for the right to search the data base and retrieve from it information or data responsive to the search request -- would infringe an existing copyright by retrieving the entire data base and marketing an exact duplicate in competition with the copyright proprietor. Such activity would beyond question be unauthorized copying in violation of a valid copyright. Purchasing access to information contained in a data base no more entitles one to make and employ copies for commercial purposes than would purchasing a copy of a copyrighted directory entitle one to produce and disseminate copies of the directory.

Two complications arise in attempting to define the scope of protection in a computerized data base. First, such works are not static; rather, they are constantly being updated by the addition of current data and the deletion of that determined obsolete. Second, the question as to what rights a copyright proprietor has in extracts of information retrieved pursuant to an authorized search of the data base must be addressed. Provisions applicable to both issues are found in the text and legislative reports of the new law.

The dynamic process by which a data base changes need not affect the entitlement of the data base to copyright protection. This process raises two concerns: 1) that deposit of a new embodiment of the data base to reflect every modification of the data therein contained would be both extremely expensive for the proprietor and cumbersome for the Library of Congress; and 2) that a proprietor, by virtue of the constant updating of the data base, could claim copyright in the work in perpetuity, in disregard of the "limited times" provision of the Constitution and the statutory term of 75 years applicable

to data bases under the new statute. Neither of these concerns need cause serious problems.

The deposit requirement should prove no bar to providing effective copyright protection for dynamic data bases. Deposit is not a precondition to copyright under the new law. Sections 407(c) and 408(c) of the new copyright atatute authorize the Register of Copyrights to exempt categories of material from the deposit requirements by regulation, or to require alternative forms of deposit. Computer data bases seem well-suited for this exemption, for the deposit of an identifying form would achieve the statutory purpose of "providing a satisfactory archival record of a work without imposing practical or financial hardships on the depositor." 13/ Nor would a dynamic data base necessarily obtain protection for a longer period than constitutionally or legislatively authorized, any more than would a telephone directory be given perpetual protection by virtue of its being updated annually. The proprietor of a data base would have to register anew periodically for copyright in such work, just as the proprietor of a telephone directory obtains copyright in new editions of a work periodically appearing.

Similar also to a telephone directory, copyright in a dynamic data base protects no individual datum, but only the systematized form in which the data is presented. The use of one item retrieved from such a work -- be it an address, a chemical formula, or a citation to an article -- would not under reasonable circumstances merit the attention of the copyright proprietor. Nor would it conceivably constitute infringement of copyright.

13/ Section 407, P.L. 94-553.

The retrieval and reduplication of any substantial portion of a data base, whether or not the individual data are in the public domain, would likely constitute a duplication of the copyrighted element of a data base, and would be an infringement. In any event, the issue of how much is enough to constitute a copyright violation would likely entail analysis on a case-by-case basis with considerations of fair use bearing on whether the unauthorized copying of a limited protion of a data base would be held non-infringing. The Commission could recommend in its Report, that fair use would have very limited force when an unauthorized copy of a data base was made for primarily commercial use. Only when information of substantial amount were extracted and duplicated for redistribution would serious problems exist, raising concerns about the enforcement of proprietary rights.

It appears that adequate legal protection for proprietary rights in extracts from data bases exists under tranditional copyright principles as expressed in the new law, supplemented by still-available relief under common law principles of unfair competition. The unauthorized taking of substantial segments of a copyrighted data base should be considered infringing, consistent with case law developed from infringement of copyright in various forms of directories.[14/] In addition, common law principles of misappropriation, which according to the legislative reports accompanying the new law are not preempted with regard to computer data bases,[15/] are available to enforce proprietary rights in these works.

14/ See, e.g., Leon v. Pacific Tel. & Tel., 91 F. 2nd 484 (9th Cir. 1937) and Jeweler's Circular Pub. Co. v. Keystone Pub. Co., 281 F. 83 (2d Cir.), cert. denied, 259 U.S. 581 (1922), aff'g 294 F. 932 (S.D.N.Y. 1921).

15/ H.R. 94-1476, supra note 3 at 132.

III. Publication

"Publication" is defined in section 101 of the new law as follows:

> "the distribution of copies or phonorecords of a work to the public by sale or other transfer of ownership, or by rental, lease, or lending. The offering to distribute copies or phonorecords to a group of persons for purposes of further distribution, public performance, or public display, constitutes publication. A public performance or display of a work does not of itself constitute publication.

According to sections 401 and 407 of the new law, after publication the copyright owner is required to place copyright notice upon all publicly distributed copies of a work, and to deposit for the Library of Congress two copies of the work. If a proprietor wishes also to register the work in accordance with section 408, the deposit required by section 407 must be accompanied by the prescribed registration application and fee. While the failure to deposit copies will not result in forfeiture of copyright, the failure to place notice on published copies may.[16/] Accordingly, it is of considerable importance to know what acts constitute publication of any copyrighted work. Computerized data bases are no exception.

16/ Under the new law, the most significant effect of the act of publication is the requirement that copyright notice be affixed to all copies of the work distributed thereafter. Omission of notice may result, in accord with the provisions contained in Section 405, in the foreiture of copyright. Section 405 of the Act of 1976 provides that omission of notice will not invalidate copyright if notice is omitted from a relatively small number of publicly distributed copies, if the work is registered within 5 years of publication and reasonable efforts are made to add notice to pubicly distributed copies, or if omission of notice violates terms set by the proprietor for authorizing public distribution of copies of the work. Section 406 deals with errors in contents of the notice with like flexibility. The failure to include notice may, at least temporarily deny the proprietor his full rights in a copyrighted work, i.e., to prevent and collect damages for unauthorized copying.

The definition cited above, and further discussed in the legislative reports accompanying the law, provides a reasonably clear benchmark for determining when a data base used in conjunction with an automated storage and retrieval system, i.e., a computer, is "published" for the purposes of the copyright law. The House Committee Report thoroughly discusses the concept of publication in the context of considering the duration of copyright under the new bill. It states that:

> "Under the definition in section 101, a work is 'published' if one or more copies or phonorecords embodying it are distributed to the public -- that is, generally to persons under no explicit or implicit restrictions with respect to disclosure of its contents -- without regard to the manner in which the copies or phonorecords changed hands. The definition . . . makes plain that any form of dissemination in which a material object does not change hands -- performance or displays on television, for example -- is not a publication no matter how many people are exposed to the work. On the other hand, the definition also makes clear that, when copies or phonorecords are offered to a group of wholesalers, broadcasters, motion pictures, [sic], etc., publication takes place if the purpose is 'further distribution, public performance, or public display.'" 17/

Accordingly, a data base proprietor could, by display alone, make the data base available to users, without having published the data base. The same would be true where the proprietor leased a tape containing the data base directly to a user and placed that user under explicit restrictions prohibiting disclosure or transfer. Under these circumstances, the failure to place copyright notice on the data base, or to register with the Copyright Office, would jeopardize no rights the proprietor might have. If, however,

---

17/ H.R. Rep. No. 1476, 94th Cong., 2nd Sess. 138 (1976).

the proprietor authorized transferees to distribute copies or make available displays of the data base, publication would be accomplished and the notice and registration requirements of the law would take affect. Many data bases are marketed in exactly this way, with the proprietor authorizing the broker to distribute or display extracts from the data base.

Certain consequences flow from the publication of any work. Publication of a work activates the requirement of deposit under section 407, and a proprietor might choose not to publish and thereby avoid the need to affix notice to all copies and deposit two copies for the Library of Congress. The doctrine of fair use may be applied more narrowly to unpublished than to published works. The Senate Report accompanying the new law indicates that "The applicicability of the fair use doctrine to unpublished works is narrowly limited since, although the work in unavailable, this is the result of a deliberate choice on the part of the copyright owner."[18/] Accordingly, the proprietor of a work may have somewhat greater rights in unpublished as opposed to published works.

Certain remedies for infringements may be made available to one who publishes and registers a work which would be denied to the proprietor of an unpublished, unregistered work under the provisions of section 412 of the Act of 1976. One who successfully prosecutes a copyright infringement action may be entitled, under section 504 of the new law, to an award of statutory damages in spite of an inability to prove actual damages. The proprietor may also be entitled to an award of attorney's fees under the provisions of section 505. Section 412 provides that the proprietor of copyright in a work neither published nor registered at the time of the infringement is not entitled to these remedies; the proprietor of a

---

18/ S. Rep. No. 473, 94th Cong., 1st Sess. 64 (1975).

published work, however, may register the work within three months after publication without forfeiting these remedies for infringing acts occurring after publication. While the key factor in determining the availability of these remedies is registration, there exists the three-month grace period after publication for registering copyright, during which period the lack of registration will not preclude availability of statutory damages and attorney's fees for infringements then occurring. No such grace period exists for registering works which are unpublished. Consistent with this thrust of the new law, the Commission's recommendations should encourage proprietors of data bases to publish and register their works and create a public record of the information available through their proprietary works.

IV. Recommendations

As previously discussed, the new law appears to deal with many of the questions raised in relation to copyright protection and automated data bases. We believe, for example, that the questions relating to whether or not input of a data base is an exclusive right included within copyright, and to the scope of protection to be provided a data base by copyright, can and should be answered in accord with the provisions already enacted in the new copyright law. Because data bases are, in some sense, unique, we believe the inclusion of appropriate language in CONTU's final report, consistent with that employed in the contents of this memorandum, will provide an appropriate interpretive aid to the statute.

The Subcommittee further believes that appropriate registration and deposit requirements should be adopted by the Register of Copyrights consistent with the statutory discretion vested in that official, to permit and encourage the

registration and periodic updating of identifying material rather than actual duplicate copies of data bases. Reasons justifying such action by the Register are found in the body of this memorandum. There appears no reason to tailor any notice requirements specifically to computer readable works: general principles contained in the new law seem adequate without being particularly burdensome. Notice appearing on the initial display of any extract or extracts obtained from the data base pursuant to a search should comply with the intent of the statutory notice requirement. Copyright notice can easily be included on the initial display extracted from a data base, and human readable notice can also appear on the packaging. This could be explained in CONTU's final report. 19/

Finally, we recommend the deletion of Section 117 of the new law, as was apparently the legislative intent upon completion of CONTU's work. Whether or not CONTU's recommendations are adopted by Congress, it would be anomalous and undesirable, as well as perhaps meaningless, to continue to include reference to pre-revision bill law within a new statutory enactment intended to be a complete codification of the law.

19/ The Copyright Office is currently drafting regulations to take effect upon the effective date of the new law. The CONTU staff has been informed that, as a result of the issues of data base deposit and notice requirements having been raised by CONTU, these issues will be the subject of proposed regulations presently being drafted.

STATEMENT OF COMMISSIONER GEORGE D. CARY

Although I concur in the overall provisions of the draft report of the Data Base Subcommittee, I should like to note a matter which may be related to the deposit of Data Bases as an element in the registration process, and which bears further consideration.

While I agree with so much of the proposed recommendation appearing on pp. 15-16 of the draft report dated 18 April 1977 that urges the Register of Copyrights to adopt appropriate regulations "to permit and encourage the registration and updating of identifying material," I have some misgivings to express agreement with that portion of the recommendation which implies that in no case should the Data Base itself be deposited.

Unless and until the Copyright Office possesses the necessary equipment which could cause the electronic signals on a tape, disc, etc. to be "perceived, reproduced or otherwise communicated either directly or with the aid of a machine or device," in a manner adequate to serve the needs of its patrons, it may serve no useful public purpose to require the deposit of a complete Data Base in each and every case. But this is not to say that in every situation "identifying material" may always be suitable or satisfactory. It is to these possibilities that I believe the door to deposit should not be closed.

Again, the Commission should be more specific with respect to the contents of the "identifying material" in lieu of copies which it is proposed to request the Register of Copyrights to embody in a regulation. The public interest is to be served, it seems to me, only if the "identifying material" contains sufficient information to indicate exactly what specific

matter is being made the subject of a particular copyright registration. In short, it should serve as an abstract of the Data Base itself.

For example, in the case of a Data Base which is continually being updated by the minute or hour, it would seem to be most unsatisfactory to one searching the records of the Copyright Office to find "identifying material" which merely specified that the particular claim of copyright covered "additions and revisions."

The "identifying material," it seems to me, must be meaningful and complete enough to direct one searching the records to an understanding of the content of the particular copyright claim. In some situations, it may be that only a reading out of the Data Base itself may be able to furnish the desired information. Hence my reluctance to recommend the acceptance by the Copyright Office of "identifying material" in all circumstances.